图书情报工作
第 65 卷 第 19 期 2021 年 10 月

ChinaXiv合作期刊

基于深度学习的技术机会预测研究*
——以新能源汽车为例

桂美增[1] 许学国[2]
[1] 浙江财经大学会计学院 杭州 310018 [2] 上海大学管理学院 上海 200444

摘 要：[目的/意义]技术机会预测有利于国家和企业管理者识别技术未来的发展方向，从而调整技术发展战略，在技术竞争中占据优势地位。[方法/过程]提出一种基于深度学习的技术机会预测方法。首先运用 AP(affinity propagation)聚类算法实现对技术领域的主题划分。其次运用 Doc2Vec 算法计算出各技术领域专利文本相似度情况，进而识别出具有发展潜力的技术领域。再次采用生成式拓扑映射(generative topographic mapping, GTM)算法对发展潜力技术领域绘制专利地图，通过 GTM 逆向映射获得技术机会。最后，构建基于深度学习的链接预测模型，对识别出的技术机会进行链接预测，从而获得高发展概率的技术机会。[结果/结论]使用新能源汽车专利数据对方法的有效性进行验证，结果显示基于深度学习的链接预测模型的预测准确率、召回率和 F1 值均优于其他预测模型，并对新能源汽车的技术机会进行预测。

关键词：技术机会预测 深度学习 链接预测 生成式拓扑映射
分类号：C931
DOI：10.13266/j.issn.0252-3116.2021.19.013

chinaXiv:202304.00471v1

随着全球科技创新活动的日益活跃，技术创新在经济发展中的作用愈加重要。加快现代工业技术体系建设，有利于我国经济高质量发展。面向“中国制造2025”，科学技术创新如何支持与引领现代工业技术系统的发展，如何实现质量、效率与动力变革的关键是科学推动新技术、新产业茁壮成长。在当前知识经济环境的背景下基于社会发展需求，分析科技发展走向，精确预测技术机会，把握技术发展方向，对于国家和企业都有重要意义。近年来随着 5G、大数据、人工智能与深度学习等信息技术的不断发展，基于数据驱动的决策、战略与预测研究逐渐成为各国关注的焦点。在此背景下运用深度学习等方法，准确的对技术机会进行预测，能够帮助企业集中优势资源攻克重点领域，打破技术桎梏，为现代产业技术体系构建编制路线图，进而帮助企业在未来国际竞争中占据有利地势。

1 研究现状

1.1 技术机会分析

技术机会分析是探讨和评估技术未来发展中存在的风险与机遇，是技术预测的一种[1]。其目的是预测技术演变的未来情景，帮助企业预测新兴技术的创新机会。该过程需要消耗专家大量的时间进行技术分析、评估与预测[2]。技术机会分析的相关定义最初是由斯坦福的 P. Schwartz 教授于 1974 年提出[3]。P. Schwartz 教授认为，技术创新机会可以为产业创新带来新的发展机遇，是可以改变产业与技术变革的重要技术。1995 年美国学者 A. L. Porter 提出了技术机会分析概念[4]，他认为技术机会分析是指在已有技术的基础上采用各种分析方法对技术未来创新可能进行识别分析。依据技术机会的定义可以发现，技术机会源于科技的不断发展过程。因此，可以将其定义为使用科学的方法对技术信息进行技术演化分析和预测的过程。根据以往的研究，技术机会分析可以分为 3 个方

* 本文系国家自然科学基金青年项目“关键利益相关者视角下新兴产业创新政策作用机制与仿真优化：以新能源汽车为例”(项目编号：71704101)研究成果之一。

作者简介：桂美增(ORCID：0000-0003-4810-8996)，讲师，博士；许学国(ORCID：0000-0001-9898-3233)，教授，博士，通讯作者，Email：xuxueguo@163.com。

收稿日期：2021-03-30 修回日期：2021-07-29 本文起止页码：130-141 本文责任编辑：徐健

chinaXiv:202304.00411v1

面[5-6]:未来技术趋势、技术创新组合和关键技术。其中,技术创新组合常称为技术空白点,是指未来具有较强发展潜力但现在仍未出现的技术组合。这些技术组合可能是未来技术创新的关键,是科技工作者与研发人员关注的焦点,因此笔者主要针对技术机会分析中的技术创新组合进行研究。

早期学者常采用德尔菲法进行技术机会分析。随着数据科学不断发展,逐渐有学者从定量分析角度对技术机会进行分析,通过文本分析、文献计量和数据挖掘等方法对科技文献信息进行有效挖掘进而实现技术机会探索。目前,技术机会分析研究方法主要包括运用专利属性数据进行分析、文本分析和技术空白点。专利属性数据主要是指表征专利原始属性的数据,这些数据均包含在专利公示信息中,如专利的国际分类号(IPC)、专利引用详情、专利申请日期与公示日期、专利权人等。H. G. Woo 等结合专利关键词和引用数据实现了技术研发的机会识别[7];S. H. Chang 等基于IPC 分类号运用专利技术网络分析方法预测了人工智能技术的技术创新机会[8];S. Y. Li 等基于专利基础信息构建了专利组合模型,进而实现了技术创新机会识别[9];Y. Tang 等基于 IPC 分类号构建专利技术共现网络,运用社团发现算法实现了 3D 打印领域的技术机会分析[10];Y. K. Ji 等基于专利 IPC 分类号计算技术相似性指标,运用相对技术优势法和国际商业潜力评价方法实现了汽车自动驾驶相关技术的机会分析[11]。

文本分析可以有效分析专利文件文本内容,这些内容对于理解技术文本中的技术特征非常关键。M. Y. Wang 等综合运用文本分析与 Orclus 聚类算法分析了微藻生物燃料的专利数据和论文数据,进而探索出该技术的潜在技术机会[12];P. Musyuni 等采用专利数据对以往冠状病毒的相关研究技术专利文本进行挖掘,预测未来发展趋势,结合预测结果识别出 Covid-19 诊断、治疗与防御措施等领域存在的技术创新机会[13];L. J. Feng 等对专利文本信息进行挖掘,采用 Word2Vec 与聚类分析相结合的方法实现了煤气层抽采技术的创新机会识别[14];X. Zhou 等基于专利文本挖掘与专利网络分析方法对人工智能相关技术进行研究,运用社区挖掘算法对重组技术的发展潜力进行预测,从而实现了人工智能领域技术创新机会识别[15];J. M. Gerken 等从语义分析的角度提出了一种语义新颖性计量方法,从而识别了汽车工业的技术创新机会[16]。

技术空白点具体是指在现有技术组合中并未存在但未来极有可能出现的技术组合。现阶段学者多采用专利地图的方法探寻技术空白点。在专利地图中,空白区域即为技术空白点。因为专利地图是由技术信息构成的数据,通过算法将原始数据从高维数据空间映射到低维正则网格上,空白网格表示该点并没有与之对应的技术组合,即为技术空白点。目前,形成专利地图常用的算法有主成分分析法[17]和自组织映射方法[18]。这两种方法均能绘制出客观性较强的专利地图,从而实现了技术空白点的识别。但在对技术空白点的解读方面,这两种算法均采用专家意见的方式,导致最终识别结果往往受到客观现实的挑战。针对这一缺陷,C. Son 等提出了一种具有逆向映射能力的技术机会分析方法——生成式拓扑映射,由于该方法能够采用逆向映射的方法获得技术空白点所表征的技术组合,因此弥补了以往方法在技术机会识别时客观性不足的缺陷[6]。我国学者吴菲菲等运用生成式拓扑映射的方法对陆地无线接入技术标注的空白点进行识别,结果验证了该方法可以很好应用于技术创新机会识别领域[19];许学国等采用生成式拓扑映射方法,从多维数据入手对专利地图的技术空白点进行分析,最终实现了技术创新机会的有效识别[20]。

1.2 链接预测

链接预测是通过分析已知的网络结构,结合其他节点的潜在信息,对网络中尚未链接的节点之间是否存在链接的可能性进行评估[21]。D. Liben 等通过对网络拓扑信息分析实现了网络节点之间的链接预测[22]。但在网络分析过程中常常存在数据不全或缺失的问题,从而导致了所形成的网络中存在部分虚假链接与未知链接。链接预测通过分析网络节点属性与链接状况能够很好地识别出网络中存在的虚假链接,并且可以对网络节点之间可能出现的链接进行判别。C. Lee 等将链接预测应用于技术网络中,通过技术节点间的链接预测,实现了技术创新机会预测[23]。

目前,学者常采用马尔科夫链、统计模型、机器学习和深度学习等方法对网络节点之间的潜在关系进行预测。J. Leskovec 等采用逻辑回归模型对社交网络中节点之间的潜在链接状况进行研究,进而识别出社交网络中的潜在好友[24];黄璐等基于链接预测方法构建了新兴技术发展识别体系,并对钙钛矿材料领域的新兴技术主题进行识别[25]。深度学习是机器学习的一个重要分支领域,是从数据中进行学习的一种新方法,现代深度学习结构通常包含数十或上百的网络层级结构,这些网络层全部都从训练数据中进行学习。因此

ChinaXiv合作期刊

相较于以往研究方法深度学习通常能够显著提升预测准确度[26]。J. Chen 等基于深度学习提出了一种 encoder-LSTM-decoder(E-LSTM-D)的预测模型对网络中的动态链接进行预测,结果表明该模型的预测结果优于传统的动态网络链接预测[27];L. Rui 等提出了一种基于堆栈降噪自编码器的链接预测模型对工业物联网移动设备的链接状况进行预测,通过历史链接信息预测下一时刻设备的链接状况[28]。随着计算机科学的不断发展,逐渐有学者从新的视角出发对链接预测进行研究。

B. Yoon 等学者发现如果将链接预测问题看作一个有监督的分类问题,则可以使用文本分类算法通过对已有知识的训练学习,实现对未来数据中的链接状态的预测。根据这一思想 B. Yoon 等采用文本分类模型支持向量机对核聚变技术、净水技术及 3D 打印技术进行链接预测,结果表明该方法在各技术领域的预测准确度均优于传统链接预测方法,预测准确率更高[29]。S. Oh 等依据专利合作信息构建了基于合作专利分类的链接预测模型,从而实现了新产品研发方向预测[30]。

可见,现阶段国内外学者对技术机会分析、链接预测已进行深入研究,但现有研究仍存在一定局限性。采用专利地图等方法进行机会识别时,许多学者侧重于寻找专利地图中的空白区域作为技术机会,并未对空白区域的真正含义进行研究。由于在专利地图中,并非所有的技术空白点均为技术机会,其中一些技术空白点由于低关联性而无法发展[29],由此可见通过技术空白点识别出的技术机会可能包含无法发展的低关联性技术机会。因此需要对识别出的技术机会进行分析,从而获得更为准确的技术机会。此外,随着人工智能技术的快速发展,深度学习以其适用性强、准确率高的特性,逐渐有学者尝试将深度学习等方法应用于技术预测与链接预测等领域[31-32],并获得了更为准确的结果。目前较少学者采用基于深度学习的链接预测方法对技术机会预测进行研究。基于此,笔者提出一种基于深度学习的链接预测方法对专利地图中的技术空白点进行链接预测,进而剔除低关联性技术空白点,最终实现技术机会预测。

2 研究框架和分析方法

2.1 研究框架

笔者提出的基于深度学习的技术机会预测分析框架(见图 1)主要包括:首先对专利数据进行收集与预处理,其次运用 apriori 算法剔除噪声信息进而构建技术相似度矩阵,然后运用 AP 算法进行聚类分析从而识别出主要技术领域。再次根据 AP 聚类分析结果采用 Doc2Vec 算法计算出识别出技术领域内所包含专利的文本相似度情况,通过该方法从整体上了解各技术领域的技术新颖性,从而筛选出具有发展潜力的技术领域[33]。并采用 GTM 绘制具有发展潜力技术领域的专利地图进而获得技术空白点,通过 GTM 逆向映射的方法获得各技术空白点所表征的技术组合。最后构建基于深度学习的链接预测模型对 GTM 识别出的技术组合与已有技术组合进行链接预测,从而挖掘出高发展概率的技术组合,最终实现技术机会预测。

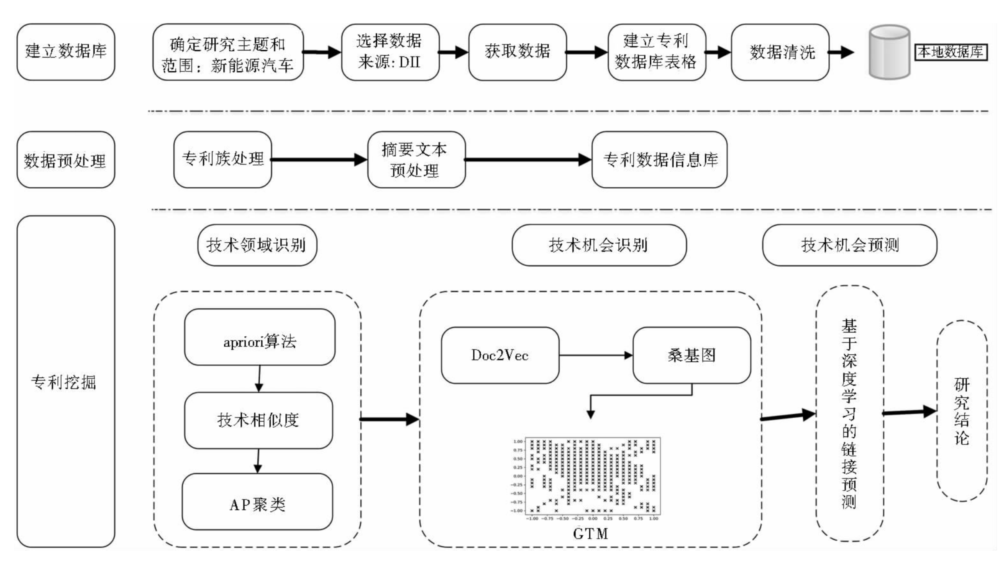


图 1 基于深度学习的技术机会预测分析框架

chinaXiv:202304.00411v1

2.2 基于 AP 聚类算法的技术领域识别

2.2.1 Affinity Propagation 算法

AP 聚类算法最初是 B. J. Frey[34]等于 2007 年在 *Science* 期刊上发表的研究成果。该算法具有不用规定聚类数目的优势，算法在迭代过程中自动获得最优聚类中心[35]。AP 聚类算法出现以后大量学者对该算法的性能与效果进行分析，R. Guan 等将 AP 聚类算法应用于半监督文本聚类，研究发现该算法的聚类结果优于传统聚类算法的结果[36]。国显达等对比分析了 AP 聚类与 K-means 算法，结果显示 AP 聚类算法性能优于 K-means 算法[37]。采用 AP 聚类算法进行聚类分析，能够克服以往聚类算法聚类结果不稳定、信息传播不完整及层次划分不合理的缺点[38]，因此笔者使用 AP 聚类算法对研究主题进行挖掘，进而获得更加客观、准确的技术领域。

2.2.2 技术相似性

由于 AP 聚类算法的输入数据为技术相似矩阵，因此首先需要对技术距离或技术相似性进行测量。笔者采用李勇敢提出的"相对技术相似度"方法，对技术距离进行测量[39]。该方法的具体过程为：A、B 为两个不同领域的技术，技术 A 相对于 B 的技术关联度（技术相似度）$R_{A\rightarrow B}$可表示为：

$$R_{A\rightarrow B}=\frac{n_{A\cap B}}{n_A}\times 100\% \qquad \text{公式(1)}$$

其中，$n_{A\cap B}$表示技术 A 与 B 的专利分类号共同出现的专利数量，n_A 表示技术 A 的专利数量。计算结果在 0－1 之间，结果越接近于 1 表示技术 A 与技术 B 越相似，反之，则越小。采用该方法构造的技术关联矩阵为非对称矩阵，矩阵中的数值为行技术对列技术的技术相似度。由于 AP 聚类算法的输入数据可以是非对称矩阵，因此笔者通过该方法构建专利技术相似矩阵作为 AP 聚类算法的输入数据，通过技术之间相似性特征进行聚类分析从而获得主要的技术领域。

2.3 基于 Doc2Vec-GTM 技术机会识别

2.3.1 基于 Doc2Vec 主题相似度测量

2013 年，Google 开源了 Word2vec 算法，该算法通过神经网络实现了词语向量化。T. Mikolov 等在 2014 年提出一种更为新颖的算法 Doc2Vec，该算法可将句子或段落直接转化为固定维度向量[40]。Doc2Vec 是在 Word2Vec 的基础上提出的一种用于测量长文本向量的深度学习算法。Doc2Vec 算法在进行文本数据训练时可以选择两种模型进行训练，分别为：DM（Distributed Memory Model）和 DBOW（Distributed Bag of Words）[41]。在训练过程中，DBOW 模式只需存储 Softmax 参数，而 DM 模式还需存储单词的词向量，训练示意图如图 2 和图 3 所示：

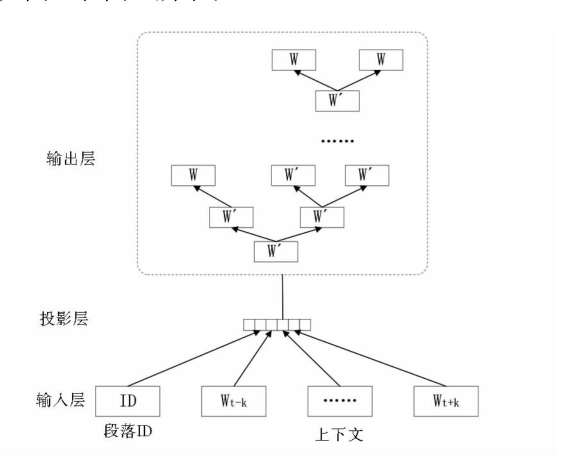


图 2 DM 模式结构

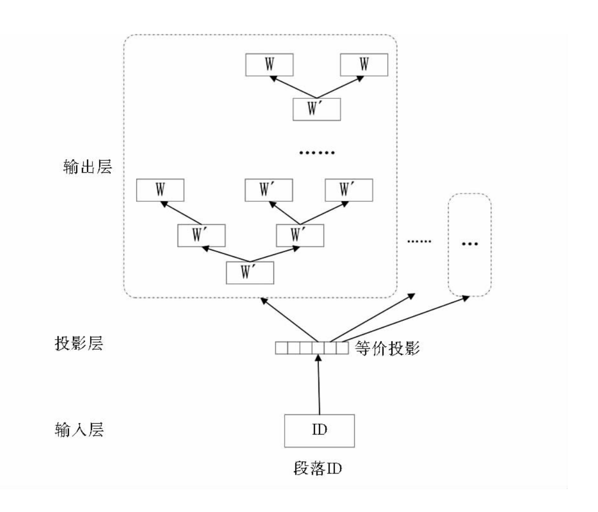


图 3 DBOW 模式结构

实践表明，Doc2Vec 在文本相似度[42]、文本分类[43]等方面表现出良好的效果。因此，笔者将 Doc2Vec 算法应用到专利文本内容的相似度测量当中，借助 Doc2Vec 算法可以保留文本上下文语境信息的优势，实现专利摘要文本的向量化描述，通过计算识别出的各技术领域内所包含专利文本相似度信息，实现从整体上了解各技术领域的技术新颖性信息，从而获得具有发展潜力的技术领域。

2.3.2 基于专利地图的技术空白点识别

随着技术竞争的日趋激烈，技术空白点识别已经成为各国技术机会挖掘的有效工具[44]。目前常用算法主要包含主成分分析法（principal component analysis，PCA），自组织映射网络（self organizing map，SOM）和生成式拓扑映射 3 种。

采用技术空白点进行技术机会识别，关键是绘制

chinaXiv:202304.00471v1

ChinaXiv合作期刊

出客观的专利地图并对技术空白点进行科学解读。PCA、SOM 和 GTM 均能绘制出客观科学的专利地图，但在技术空白点解读方面，PCA 与 SOM 均需要借助专业人员进行主观判断，进而分析出技术潜在发展方向。而 GTM 具有逆向映射专利地图的能力（如图 4 所示），通过逆向映射可以直接获得技术空白点所表征的技术组合。基于 GTM 形成的专利地图在解释专利空白方面，无论是效率还是效果都优于另外两种方法[45]。因此，本文采用 GTM 对识别出的技术领域绘制专利地图，从而获得技术空白点，通过 GTM 逆向映射获得各技术空白点所表征的技术组合。

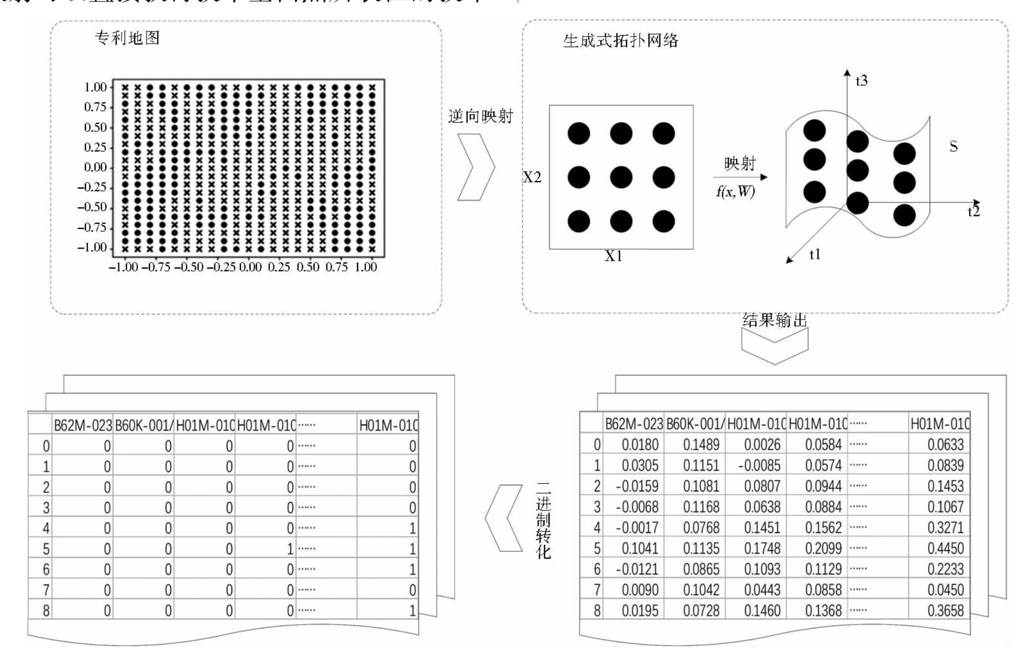


图 4　GTM 逆向映射示意

2.4　基于深度学习的技术机会预测

在专利地图中，并非所有的技术空白点均为技术机会，其中一些技术空白点由于与现有技术的低关联性而无法发展。从技术持续创新发展的角度来看，若某技术空白点与最新申请的专利具有高度关联性，那么它具有较大机会被开发实现，因为新技术的研发成功通常是在已有技术的基础上不断创新发展所形成的。因此需要对专利地图中识别出的技术空白点进行分析，进而识别出真正的技术机会。笔者采用链接预测的方法对技术空白点进行技术评估，通过预测技术空白点与已有技术组合的潜在链接情况，进而探索出最具发展潜力的技术机会。

常用的链接预测方法有基于节点邻域分析、路径分析方法和分类方法。由于采用 GTM 识别出的技术空白点与已有专利均未存在联系，采用基于节点邻域分析方法无法实现链接预测。因此笔者采用分类方法对技术空白点与已有技术信息的关联性进行预测。文本分类方法在机器学习领域属于监督学习，采用监督学习对已有的技术信息进行学习，获得技术之间的潜在关系，进而实现对技术空白点与现有技术之间的链接预测。目前文本分类领域深度学习以其预测准确率高、适用性强等特点已广泛应用于各研究领域，并取得了较好的预测结果。因此为了提高链接预测准确率，笔者尝试构建基于深度学习的链接预测模型对 GTM 识别出的技术机会与已有技术组合进行链接预测，从而挖掘出高发展概率的技术机会。

为了采用深度学习对技术之间的链接关系进行预测，首先需要构建训练样本与测试样本。由于技术创新发展过程在科技文献引用信息中有所体现，孙冰等构建专利引文网络，从而对手机芯片的核心专利进行识别，进而对该专利技术的扩散路径进行分析[46]；柴庆凤等基于专利引用信息构建知识流动网络，采用支持向量机和多元线性回归方法构建链接预测模型，从而实现了技术链接机会预测[47]；C. Lee 等在专利引用信息的基础上结合专利基础信息提出了一种基于机器学习的关键新兴技术识别模型，通过该模型实现了制药领域关键新兴技术的识别[48]。因此，笔者根据专利之间的引用关系来表征专利技术持续发展过程，从而构建链接预测的样本数据，数据构建的详细过程如表 1 所示：

chinaXiv:202304.00471v1

表 1　训练数据和测试数据

类别	施引专利(CA)	被引专利(CB)	IPC(CA)	IPC(CB)	链接
训练数据	US2014232233-A1	US7999431-B2	(…,…,…)	(…,…,…)	1
	US5761028-A	US3916433-A	(…,…,…)	(…,…,…)	1
	WO2018026850-A1	EP2846380-A1	(…,…,…)	(…,…,…)	1
	…	…	…	…	…
	JP2005017156-A	EP1288051-A2	(…,…,…)	(…,…,…)	0
	JP2017123294-A	CN106904079-A	(…,…,…)	(…,…,…)	0
	TR201612616-A	CN106560362-A	(…,…,…)	(…,…,…)	0
测试数据	CA2820449-A1	EP2677177-A1	(…,…,…)	(…,…,…)	1
	CN108116246-A	CN205407309-U	(…,…,…)	(…,…,…)	1
	…	…	…	…	…
	CN108119308-A	KR2018058587-A	(…,…,…)	(…,…,…)	0
	JP2018085790-A	CN205811017-U	(…,…,…)	(…,…,…)	0

输出数据为专利之间的链接情况。若两项专利存在(不存在)引用关系,则归类为"链接(不链接)"。数据处理过程采用 Python 语言编写程序,首先提取专利数据的 PN 信息获得专利数据的专利号作为专利数据的标识信息。其次,提取专利数据中的 CP 信息获得专利数据的引用信息。最后,根据专利数据的标识信息与引用信息确定相关数据的链接情况,若两项专利存在(不存在)引用关系,则归类为"链接(不链接)"。进而获得原始数据样本,并按照 4:1 的比例随机筛选训练数据与测试数据。

笔者采用链接预测的方法对通过 GTM 逆向映射识别出的技术空白点进行评估,通过预测技术空白点与已有技术组合的潜在链接情况,进而探索出最具发展潜力的技术机会。构建的基于深度学习的链接预测模型框架如图 5 所示:

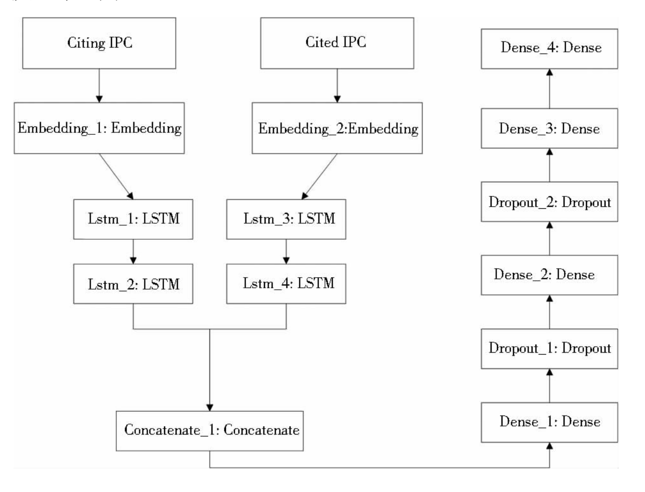


图 5　基于深度学习的链接预测模型

基于深度学习的链接预测模型的输入数据为施引专利的 IPC 和被引专利的 IPC,通过词嵌入层(Embedding)对输入数据进行词向量化处理。随后运用连接层将分别通过双层长短记忆神经网络(LSTM)处理后的输入数据进行合并。接着添加 3 个激活函数为 relu 的全连接层(Dense),并在他们之间添加 Dropout 层。最后添加激活函数为 sigmoid 的分类器。通过输入数据进行训练,获得链接预测模型。

笔者使用 Python 语言选择深度学习 Keras 库搭建基于深度学习的链接预测模型,计算后台选用"Tensorflow"。在训练过程中为了保证模型输入数据长度统一,采用 Keras 库中的"pad_sequences"函数对输入数据进行填充。此外训练过程中模型优化器选择"RMSProp",batch_size 设置为 64,由于最终节点设置判别为真和假,因此损失函数设置为"binary_crossentropy"。通过训练完成的链接预测模型能够对 GTM 识别出的技术空白点与已有技术组合信息进行链接预测,从而实现对技术空白点的机会预测。

3　实证研究

3.1　数据收集与预处理

3.1.1　数据收集

专利作为科学技术的载体,经过相关领域专家的严密审核和科学验证,能够准确地反映出技术内容,因此学者们常采用专利信息进行技术机会的挖掘。笔者以新能源汽车技术为例进行方法有效性研究,采用 TI = (New energy vehicle OR New energy automobile OR Battery Electric Vehicle OR Battery Electric automobile OR pure electric vehicle OR pure electric automobile OR hybrid electric vehicle OR hybrid electric automobile)检索式在德温特专利数据库进行检索,时间设置为 1968-2019 年,检索时间为 2020 年 8 月 11 日,共检索到

chinaXiv:202304.00471v1

ChinaXiv合作期刊

71 064项专利(包含专利族)数据。

3.1.2 数据预处理

首先根据专利公开号对专利族进行处理。然后依据专利优先权申请和日期,提取出各专利的申请日期,由于 1995 年之前新能源汽车领域的专利数量仅为 2 780项,因此笔者分时段筛选出 1995 – 1999 年、2000 – 2004 年、2005 – 2009 年、2010 – 2014 年和 2015 – 2019 年专利数据进行分析。随后,提取每个专利的国际分类号 IPC。最后提取各专利的摘要文本,并对摘要文本进行停用词剔除、词干提取等文本预处理,进而确保模型结果质量。

3.2 基于 AP 聚类算法的技术领域识别

专利是技术情报中最重要的载体之一,专利文档中对技术情报的记录准确且丰富,专利是企业参与技术竞争与发展最重要的资源之一。每个专利数据均包含了高度概括其技术范围的国际专利分类号(IPC),该分类号是由专利机构审查人员根据技术标准对专利进行标注,通过查阅相关的技术手册可以轻松查询到该分类号所表征的技术领域。并且当专利拥有两个及以上的分类号时,则表明该专利的内容涉及多个技术主题,这意味着这些分类号所表示的技术主题之间存在关系。通过对专利分类号分析可以实现技术关联性、技术差异性与技术交叉性等方面的分析。因此采用专利数据中 IPC 分类号进行技术领域识别能够进一步增强分析结果的准确性。

为了使用 AP 聚类算法进行聚类分析,首先需要构建技术相似矩阵作为输入数据。由于现阶段需要分析的数据比较庞大,传统的统计方法难以进行大规模的深度分析,所以笔者使用数据挖掘的方法以获取有价值的知识[49]。关联分析作为一种有效的数据挖掘工具,之前常用在分析客户的消费习惯、员工的组织特征等领域。现阶段已有学者运用关联规则对不同量级知识数据进行挖掘,研究表明采用关联规则方法从知识共现关系中提取技术联系的可能[50]。因此采用 apriori 算法对数据进行挖掘,可以从海量数据中快速识别出其中主要的组成部分,作为 AP 聚类分析的输入数据。通过该方法可以过滤掉大量的噪声信息,从而保证分析结果的质量[51]。笔者参考以往学者经验将 apriori 算法的最小支持度设置为 0.01,最小置信度设置为 0.3[52]。以 2015 – 2019 年为例,共识别出 65 项 IPC,根据公式(1)构建 65 * 65 的技术相似度矩阵,详细结果如表 2 所示:

表 2 技术相似性矩阵

	B60L-050/64	H01M-002/10	H01M-010/615	B60W-020/13	B60L-011/18	B62D-025/20	…	H01M-002/26	H01M-010/63	H01M-002/30	G01R-031/36	H01M-010/6551	H01M-010/6563
B60L-050/64	1.00	0.74	0.04	0.02	0.42	0.05	…	0.04	0.01	0.09	0.01	0.03	0.03
H01M-002/10	0.07	1.00	0.06	0.00	0.21	0.03	…	0.09	0.03	0.09	0.02	0.09	0.07
H01M-010/615	0.02	0.37	1.00	0.00	0.40	0.07	…	0.08	0.21	0.02	0.02	0.07	0.16
B60W-020/13	0.01	0.01	0.00	1.00	0.44	0.00	…	0.00	0.00	0.00	0.05	0.00	0.01
B60L-011/18	0.03	0.16	0.05	0.05	1.00	0.02	…	0.02	0.03	0.01	0.06	0.01	0.03
B62D-025/20	0.07	0.49	0.17	0.00	0.45	1.00	…	0.17	0.17	0.00	0.00	0.00	0.00
…	…	…	…	…	…	…	…	…	…	…	…	…	…
H01M-002/26	0.02	0.51	0.08	0.00	0.13	0.06	…	1.00	0.07	0.37	0.03	0.03	0.02
H01M-010/63	0.01	0.39	0.52	0.01	0.56	0.16	…	0.17	1.00	0.01	0.03	0.05	0.15
H01M-002/30	0.05	0.62	0.02	0.00	0.12	0.00	…	0.43	0.00	1.00	0.04	0.06	0.01
G01R-031/36	0.00	0.09	0.01	0.02	0.28	0.00	…	0.02	0.01	0.02	1.00	0.01	0.01
H01M-010/6551	0.02	0.84	0.12	0.00	0.16	0.00	…	0.04	0.03	0.08	0.02	1.00	0.17
H01M-010/6563	0.02	0.58	0.22	0.01	0.31	0.00	…	0.02	0.08	0.01	0.02	0.16	1.00

获得技术相似矩阵后,笔者使用 AP 聚类算法进行技术领域识别。AP 聚类使用的是近邻传播方法,其能够较好地解决聚类过程中对数据的主题聚类数目的确定。该算法不用提前规定结果簇的个数。笔者采用 Python 语言使用机器学习“sklearn”库中的 Affinity Propagation 算法,将各时期的技术关系矩阵输入 AP 聚类算法进行聚类,获得各时期的聚类结果。

3.3 基于 Doc2Vec 专利文本相似度测量

通过 AP 聚类算法的聚类结果可以确定每项专利所属的技术领域,为了从整体上研究各技术领域的技术新颖性,笔者采用 Doc2Vec 算法对专利摘要文本进行训练,从而计算出各领域内摘要文本的相似度状况。在实验过程中主要参数的设定为:DM = 1(即采用 DM 模式),window 设定为 6,文本向量维度设置为 50。在

chinaXiv:202304.00471v1

实验中，为提高计算的精度，笔者将 epochs 的值设定为 40，且每次 epoch 都对训练数据重新打乱。最终计算获得各技术领域内文本相似度情况，如表 3 所示：

表 3　文本相似度情况

编号	1995－1999 年	2000－2004 年	2005－2009 年	2010－2014 年	2015－2019 年
1	0.298 2	0.241 7	0.197 5	0.263 2	0.152 9
2	0.183 2	0.235 3	0.163 4	0.182 6	0.223 8
3	0.290 1	0.334 9	0.204 8	0.223 8	0.223 2
4	0.230 5	0.210 7	0.339 9	0.161 9	0.209 9
5	0.231 2	0.279 3	0.370 8	0.218 3	0.326 4
6	0.203 9	0.285 9	0.274 8	0.320 6	0.221 8
7		0.270 2	0.301 3	0.229 1	0.171 3
8		0.190 6	0.240 4	0.249 9	0.212 4
9				0.182 2	0.214 5
10				0.353 1	0.193 5
11					0.201 4

表 3 展示了 1995－1999 年、2000－2004 年、2005－2009 年、2010－2014 年和 2015－2019 年这 5 个时期通过 AP 聚类识别出的各技术领域的组内文本相似度情况。为了更加清晰观察不同时期各技术领域的变化趋势，笔者绘制了 1995－2019 年技术变化桑基图，见图 6。

如图 6 所示，其中每列代表 5 个不同时期采用 AP 聚类算法识别出的主要技术领域，为了便于观察将每列数据统一命名。例如 1995C1，该项表示 1995－1999 年这一时期簇 1 中所包含的所有 IPC，线条宽度表示簇 1 所包含的 IPC 数量。不同簇之间的连接信息为各簇之间存在相同 IPC 的情况，连线的宽度表示两簇之间具有相同 IPC 的数量。由于随着技术的不断发展，其中一些 IPC 仅存在于其中一段时期，因此无法在桑基图中呈现。结合图 6 和表 3 可以发现 1995－1999 时期

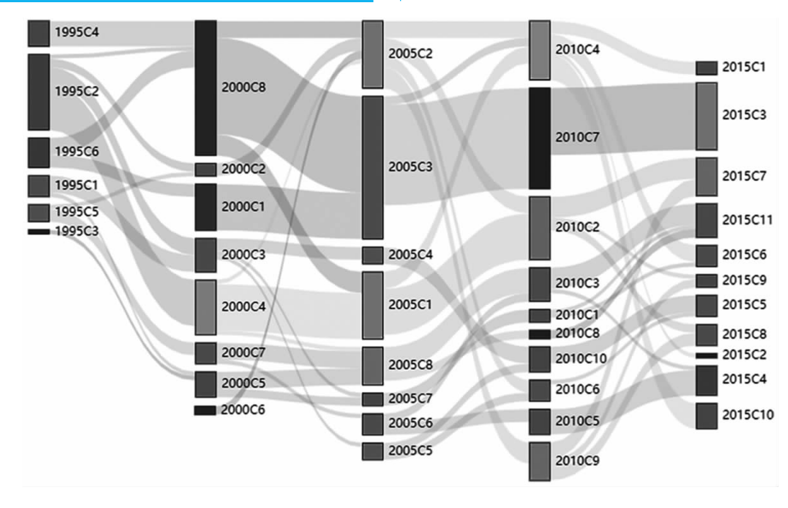


图 6　1995－2019 年新能源汽车技术变化桑基图

文本相似度最低的簇 2 最为活跃，该领域技术持续发展与其它新技术不断结合发展形成了 2000－2004 时期的簇 2、8、3 和 4 等技术领域，是 1995－1999 时期最为活跃的技术领域。除此之外 1995－1999 年文本相似度较低的簇 6 也不断发展为下一期的簇 1 和 8 等技术领域。其余时期各技术领域的技术变化趋势也呈现出相似的特征，如 2000－2004 年文本相似度较低的簇 4、簇 8 均不断发展，其中许多技术不断融合逐渐发展成为下一期新的技术领域。该结果印证了采用 AP 聚类对技术领域识别的基础上结合自然语言处理技术对技术领域文本相似度进行分析能够较好地识别出具有发展潜力的技术领域。

3.4　基于 GTM 技术机会识别

通过 AP 聚类算法和 Doc2Vec 算法实现了对新能源汽车领域不同时期的发展潜力技术领域识别。为了对发展潜力领域技术进行深入挖掘，笔者采用 GTM 对发展潜力领域进行技术空白点分析，从而获得技术机会，输入数据为所属该簇专利 IPC 数据。在 GTM 参数设置方面，笔者采用 21 * 21 维的高斯函数作为基函数，基函数的宽度值设置为 2，正则化参数为 0.001。以 2015－2019 年相似度最低的簇 1 数据为例进行分析，簇 1 专利所形成的专利地图见图 7。

在 GTM 所绘制的专利地图中，每个点代表一种专利(“×”)，空白点(“●”)是未来可能出现的技术专利，其中可能包含多项关键技术，这些技术组合可能是未来技术发展的关键创新点。为了获得各技术空白点所表征的具体技术，需要通过 GTM 逆向映射，将技术空白点逆向映射到原始数据空间，从而完成技术空白点的解析(见图 4)，可以看出识别出的技术空白点对应了表格中的每一行，表格的数据是各技术空白点经过 GTM 逆向映射后出现在原始空间对应技术下的概

chinaXiv:202304.00471v1

ChinaXiv合作期刊

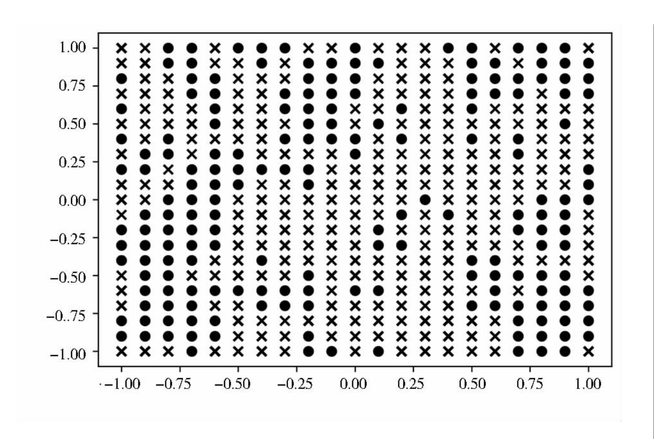


图 7　2015 – 2019 年簇 1 专利地图

率。为了获得各技术空白点所表征的技术组合，需要设置阈值对各技术出现的概率进行比较，其中大于阈值的技术均被保留下来。现阶段没有统一确定阈值的方法，以往学者常通过研究目的对阈值进行确定。通常情况下，若所取的阈值越大，识别出的技术空白点就越少。笔者借鉴以往研究成果将阈值设为 0.3 进而得到最终的技术矢量矩阵[53]，从而获得各技术空白点所表征的技术组合，部分结果见表 5。

3.5　基于深度学习的技术机会预测

为了对 GTM 识别出的空白技术组合进行深入分析，笔者采用深度学习的方法构建链接预测模型（Deep Learning Link Prediction，DLLP）对各技术空白点进行链接预测。模型的训练数据与测试数据根据专利引用信息进行筛选，共获得 11 768 项数据样本，其中存在引用关系数据 5 870 项，不存在引用关系 5 898 项。按照训练集与测试集 4:1 的比例进行划分，最终获得训练所需的训练数据 8 140 项与测试数据 3 628 项。随后分别将训练数据与测试数据放入已经设置好的基于深度学习的链接预测模型当中进行训练，从而获得最终的预测模型。此外，笔者选择常用的文本分类模型支持向量机（SVM）、决策树（DT）、贝叶斯（NB）、逻辑回归（LR）和集成学习算法 Bagging、Xgboost 进行对照实验，采用准确率、召回率与 F1 值对比分析该模型的分类效果。具体结果如表 4 所示：

表 4　模型误差统计

分类模型类别	准确率	召回率	F1 值
DLLP	0.813	0.807	0.810
DT	0.707	0.701	0.704
LR	0.670	0.667	0.668
NB	0.666	0.658	0.662
SVM	0.757	0.756	0.756
Bagging	0.779	0.779	0.779
Xgboost	0.764	0.762	0.763

根据表 4 可以发现，笔者提出的基于深度学习的链接预测模型在预测准确率、召回率及 F1 值均优于其他模型。结果表明该方法能够显著提升预测精度，其预测结果准确性更高。基于深度学习的链接预测模型能够很好地实现专利之间的链接预测。因此运用该模型对采用 GTM 获得的空白技术组合和已有专利进行链接预测分析，其中输入数据为空白技术组合的 IPC 信息与现有专利的 IPC 信息，模型输出结果即为该空白技术组合与已有专利的链接预测值。由于笔者构建的深度学习模型最后一层全连接层采用“sigmoid”激活函数，因此模型的预测结果被压缩到 0 – 1。若某项空白技术组合与已有专利具有较强的关联性，则预测结果接近 1，反之亦然。其中，每个空白技术组合与多项已有专利存在链接预测值，笔者取各项空白技术组合的最高链接预测值为该空白技术组合的预测结果，依据预测结果从而识别出具有高发展概率的技术机会。

通过预测结果可以轻松识别出具有较高发展概率的技术机会。笔者以 2015 – 2019 年簇 1 数据为例对 GTM 识别出的空白技术组合进行分析，通过对空白技术组合与已有专利信息进行链接预测，进而获得各空白技术组合的最终预测结果。其中预测值大于 0.9 的空白技术组合有 11 项，如表 5 所示：

表 5　技术机会预测结果

序号	IPC	预测值
1	B60L-053/22；B60L-011/18；H02J-007/00；B60L-053/14；H02J-050/10；B60L-053/12；H02J-007/02；	0.93
2	B60L-011/18；H02J-007/00；B60L-053/14；H02J-050/10；B60L-053/12；H02J-007/02；	0.93
3	B60L-050/60；B60L-003/00；B60L-053/22；B60L-011/18；B60L-001/00；H02J-007/00；B60L-053/14；B60L-053/20；H02M-003/335；B60L-058/22；	0.93
4	B60L-050/60；B60L-003/00；B60L-053/22；B60L-011/18；B60L-001/00；H02J-007/00；B60L-053/20；	0.93
5	B60L-050/60；B60L-003/00；B60L-053/22；B60L-011/18；B60L-001/00；H02J-007/00；B60L-053/14；B60L-053/20；H02J-007/34；H02M-003/335；	0.93
6	B60L-050/60；B60L-003/00；B60L-053/22；B60L-011/18；B60L-001/00；H02J-007/00；B60W-010/26；B60L-053/20；H02J-007/34；H02M-003/335；	0.93
7	B60L-003/00；B60L-011/18；H02J-007/00；B60R-016/033；H01M-010/48；H01M-010/42；	0.93
8	B60L-011/18；H02J-007/00；B60L-053/30；B60L-053/16；B60L-053/14；B60L-053/60；H02J-007/02；	0.92
9	B60L-050/60；B60L-003/00；B60L-053/22；B60L-011/18；B60L-001/00；H02J-007/00；B60L-053/14；B60L-053/20；G01R-031/02；H02J-007/34；H02M-003/335；	0.92
10	B60L-003/00；B60L-011/18；B60L-001/00；H02J-007/00；B60R-016/03；B60R-016/033；	0.92
11	B60L-011/18；H02J-007/00；B60L-053/30；B60L-053/16；B60L-053/14；B60L-053/60；H02J-050/10；B60L-053/12；H02J-007/02；	0.9

chinaXiv:202304.00471v1

为了深入分析识别出的技术机会,查阅 IPC 所对应的技术范围就能分析出每项高发展概率的技术机会所表示的技术领域。其中 1、2 和 11 涉及无线充电技术,4、5 涉及车辆电力逆变器,6 涉及整车控制系统,7 和 9 涉及电池监控装置,10 涉及整车集成技术。根据 2019 年、2020 年世界新能源汽车大会研究报告和相关学者研究[54]指出,未来新能源汽车研究的前沿技术及重点技术领域主要包括电动汽车无线充电技术、800 伏逆变器技术、电池动力系统、电池控制系统技术、整车控制系统和整车集成技术等领域。对比分析本文获得的技术机会预测结果可以发现,其结果基本涉及专家预测的技术领域,该结果进一步验证了该方法的有效性。

从实验结果可以发现,笔者提出的基于深度学习的技术机会预测模型能够有效实现专利技术知识流动的预测,通过对专利国际分类号的分析,能够提早识别出技术在创新融合过程中的研究方向,进而实现技术机会预测。

4 结论与不足

在当今技术快速发展的新时期,有效实现技术机会预测,有利于国家和企业管理者识别出技术的未来发展方向,从而调整发展战略,为技术竞争占据有利态势。基于此,笔者运用深度学习结合专利文本信息及专利引用关系构建了一种新的技术机会预测方法。以新能源汽车技术为例,实证检验了该方法的可行性,并对新能源汽车的技术机会进行预测。本研究为基于人工智能、大数据分析的技术机会预测提供了一种新的解决思路与方法,进而为我国其他技术领域技术机会预测提供了一定的借鉴意义。

当然,本文的研究还存在一定不足,这将成为后续的研究重点。如本文仅考虑使用专利数据对新能源汽车的主要技术领域进行技术机会预测,由于不同的文献类型代表了技术发展的不同阶段,本研究结果仅能体现技术实验阶段的实际情况,在实际决策时需要考虑其他相关维度信息[55]。此外 GTM 的参数设定比较敏感,如果参数设定失误将导致整个研究的失败。

参考文献:

[1] OLSSON O. Technological opportunity and growth[J]. Journal of economic growth, 2005, 10(1): 31-53.

[2] LEE J, KIM C, SHIN J. Technology opportunity discovery to R&D planning: key technological performance analysis[EB/OL]. [2021-01-06]. https://doi.org/10.1016/j.techfore.2017.03.011.

[3] SCHWARTZ P. Technological innovation opportunities[J]. Computers and people, 1974, 23(5): 33.

[4] PORTER A L, DETAMPEL M J. Technology opportunities analysis[J]. Technological forecasting and social change, 1995, 49(3): 237-255.

[5] 潘东华,徐珂珂. 基于共词分析的技术机会分析[J]. 科研管理, 2014, 35(4): 12-19.

[6] SON C, SUH Y, JEON J, et al. Development of a GTM-based patent map for identifying patent vacuums[J]. Expert systems with applications, 2012, 39(3): 2489-2500.

[7] WOO H G, YEOM J, LEE C. Screening early stage ideas in technology development processes: a text mining and k-nearest neighbours approach using patent information[J]. Technology analysis and strategic management, 2019, 31(5): 532-545.

[8] CHANG S H. Technical trends of artificial intelligence in standard-essential patents[J]. Data technologies and applications, 2021, 55(1): 97-117.

[9] LI S Y, ZHANG X, XU H Y, et al. Measuring strategic technological strength: patent portfolio model[EB/OL]. [2021-01-06]. https://doi.org/10.1016/j.techfore.2020.120119.

[10] TANG Y, LOU X M, CHEN Z F, et al. A study on dynamic patterns of technology convergence with IPC co-occurrence-based analysis: the case of 3D printing[J]. Sustainability, 2020, 12(7): 1-26.

[11] JI Y K, ZHU X Z, ZHAO T W, et al. Revealing technology innovation, competition and cooperation of self-driving vehicles from patent perspective[EB/OL]. [2021-01-06]. https://doi.org/10.1109/ACCESS.2020.3042019.

[12] WANG M Y, FANG S C, CHANG Y H. Exploring technological opportunities by mining the gaps between science and technology: microalgal biofuels[EB/OL]. [2021-01-16]. https://doi.org/10.1016/j.techfore.2014.07.008.

[13] MUSYUNI P, AGGARWAL G, NAGPAL M, et al. A case study: analysis of patents on coronaviruses and covid-19 for technological assessment and future research[J]. Current pharmaceutical design, 2021, 27(3): 423-439.

[14] FENG L J, NIU Y X, LIU Z F, et al. Discovering technology opportunity by keyword-based patent analysis: a hybrid approach of morphology analysis and usit[J]. Sustainability, 2020, 12(1): 1-35.

[15] ZHOU X, HUANG L, ZHANG Y, et al. A hybrid approach to detecting technological recombination based on text mining and patent network analysis[J]. Scientometrics, 2019, 121(2): 699-737.

[16] GERKEN J M, MOEHRLE M G. A new instrument for technology monitoring: novelty in patents measured by semantic patent analysis[J]. Scientometrics, 2012, 91(3): 645-670.

[17] LEE S, YOON B, PARK Y. An approach to discovering new technology opportunities: keyword-based patent map approach[J]. Technovation, 2009, 29(6): 481-497.

chinaXiv:202304.00411v1

ChinaXiv合作期刊

[18] CHOI J, HWANG Y S. Patent keyword network analysis for improving technology development efficiency[EB/OL]. [2021-01-06]. https://doi.org/10.1016/j.techfore.2013.07.004.

[19] 吴菲菲,米兰,黄鲁成. 以技术标准为导向的企业研发方向识别与评估[J]. 科学学研究, 2018, 36(10): 1837-1847.

[20] 许学国,桂美增. 基于GTM逆向映射的技术创新机会识别——以新能源汽车为例[J]. 情报理论与实践,2021,44(6):146-153,198.

[21] 翟东升,刘鹤,张杰,等. 一种基于链路预测的技术机会挖掘方法[J]. 情报学报, 2016, 35(10): 1090-1100.

[22] LIBEN D, KLEINBERG J. The link prediction problem for social networks[J]. Journal of the American Society for Information Science and technology, 2007, 58(7): 1019-1031.

[23] LEE C, KANG B, SHIN J. Novelty focused patent mapping for technology opportunity analysis[EB/OL]. [2021-01-06]. https://doi.org/10.1016/j.techfore.2014.05.010.

[24] LESKOVEC J, HUTTENLOCHER D, KLEINBERG J. Predicting positive and negative links in online social networks[C]//Proceedings of the 19th international conference on world wide web. New York: ACM Press, 2010: 641-650.

[25] 黄璐,朱一鹤,张嶷. 基于加权网络链路预测的新兴技术主题识别研究[J]. 情报学报, 2019, 38(4): 335-341.

[26] ZHOU M, WANG B, GUO S D, et al. Multi-objective prediction intervals for wind power forecast based on deep neural networks[EB/OL]. [2021-01-06]. https://doi.org/10.1016/j.ins.2020.10.034.

[27] CHEN J, ZHANG J, XU X, et al. E-LSTM-D: a deep learning framework for dynamic network link prediction[J]. IEEE transactions on systems man cybernetics-systems, 2021, 51(6): 3699-3712.

[28] RUI L, ZHU Y, GAO Z, et al. CLPM: a cooperative link prediction model for industrial internet of things using partitioned stacked denoising autoencoder[J]. IEEE transactions on industrial informatics, 2021, 17(5): 3620-3629.

[29] YOON B, MAGEE C L. Exploring technology opportunities by visualizing patent information based on generative topographic mapping and link prediction[EB/OL]. [2021-01-06]. https://doi.org/10.1016/j.techfore.2018.01.019.

[30] OH S, CHOI J, KO N, et al. Predicting product development directions for new product planning using patent classification-based link prediction[J]. Scientometrics, 2020, 125(3): 1833-1876.

[31] 许学国,桂美增. 基于深度学习的技术预测方法——以机器人技术为例[J]. 情报杂志, 2020, 39(8): 53-62.

[32] DHELIM S, NING H, AUNG N. Compath: user interest mining in heterogeneous signed social networks for internet of people[J]. IEEE Internet of things journal, 2021, 8(8): 7024-7035.

[33] KIM M, PARK Y, YOON J. Generating patent development maps for technology monitoring using semantic patent-topic analysis[EB/OL]. [2021-01-06]. https://doi.org/10.1016/j.cie.2016.06.006.

[34] FREY B J, DUECK D. Clustering by passing messages between data points[J]. Science, 2007, 315(5814): 972-976.

[35] TAHERI S, BOUYER A. Community detection in social networks using affinity propagation with adaptive similarity matrix[J]. Big data, 2020, 8(3): 189-202.

[36] GUAN R, SHI X, MARCHESE M, et al. Text clustering with seeds affinity propagation[J]. IEEE transactions on knowledge and data engineering, 2010, 23(4): 627-637.

[37] 国显达,那日萨,高欢,等. 基于Gaussian LDA的在线评论主题挖掘研究[J]. 情报学报, 2020, 39(6): 630-639.

[38] YANG Y, DORN C. Affinity propagation clustering of full-field, high-spatial-dimensional measurements for robust output-only modal identification: a proof-of-concept study[EB/OL]. [2021-02-06]. https://doi.org/10.1016/j.jsv.2020.115473.

[39] 李勇敢. 技术领域维度下相对技术关联度研究——以德温特专利数据库共类分析为例[J]. 科技进步与对策, 2017, 34(7): 146-153.

[40] LE Q, MIKOLOV T. Distributed representations of sentences and documents[C]// Proceedings of the 31st international conference on machine learning. New York: PMLR, 2014: 1188-1196.

[41] FENG S. The proximity of ideas: an analysis of patent text using machine learning[J]. Plos one, 2020, 15(7): 1-19.

[42] LU Y, XIONG X, ZHANG W, et al. Research on classification and similarity of patent citation based on deep learning[J]. Scientometrics, 2020, 123(2): 813-837.

[43] KIM D, SEO D, CHO S, et al. Multi-co-training for document classification using various document representations: TF-IDF, LDA, and doc2vec[EB/OL]. [2021-02-06]. https://doi.org/10.1016/j.ins.2018.10.006

[44] TRAPPEY A J C, CHEN P P J, TRAPPEY V C, et al. A machine learning approach for solar power technology review and patent evolution analysis[J]. Applied sciences-basel, 2019, 9(7): 1-25.

[45] BISHOP C M, SVENSEN M, WILLIAMS C K I. GTM: the generative topographic mapping[J]. Neural computation, 1998, 10(1): 215-234.

[46] 孙冰,徐晓菲,苏晓. 技术扩散主路径及核心企业的识别研究——以手机芯片专利引文网络为例[J]. 情报学报, 2019, 38(2): 201-208.

[47] 柴庆凤,翟东升,蔡力伟,等. 基于专利网络链接模型的技术链接机会预测方法研究[J]. 情报理论与实践, 2020, 43(12): 111-119.

[48] LEE C, KWON O, KIM M, et al. Early identification of emerging technologies: a machine learning approach using multiple patent indicators[EB/OL]. [2021-01-06]. https://doi.org/10.1016/j.techfore.2017.10.002.

[49] SOHRABI M K, ROSHANI R. Frequent itemset mining using cellular learning automata[EB/OL]. [2021-01-06]. https://doi.org/10.1016/j.chb.2016.11.036.

chinaXiv:202304.00471v1

ChinaXiv合作期刊

[50] ALI S H. Miner for oaccr: case of medical data analysis in knowledge discovery[C]//2012 6th international conference on sciences of electronics, technologies of information and telecommunications (setit). Sousse: IEEE, 2012:962－975.
[51] DA COSTA M, DOS SANTOS L, SCHAEFER J, et al. Industry 4.0 technologies basic network identification[J]. Scientometrics, 2019, 121(2): 977－994.
[52] 贾军,魏洁云. 新兴产业核心技术早期识别方法与应用研究[J]. 科学学研究, 2018, 36(7): 1206－1214.
[53] 吴菲菲,陈明,黄鲁成. 基于 GTM 的 3D 生物打印专利技术空白点识别[J]. 情报杂志, 2015, 34(3): 58－64.
[54] 郭本海,陆文茜,王涵,等. 基于关键技术链的新能源汽车产业政策分解及政策效力测度[J]. 中国人口・资源与环境, 2019, 29(8): 76－86.
[55] MARTINO J P. A review of selected recent advances in technological forecasting[J]. Technological forecasting and social change, 2003, 70(8): 719－733.

作者贡献说明:

桂美增:论文构思、撰写和修改;

许学国:论文构思和修改。

Approach to Technology Opportunity Prediction Based on Deep Learning: Taking the Case of New Energy Vehicles

Gui Meizeng[1] Xu Xueguo[2]

[1] Accounting college, Zhejiang University of Finance and Economics, Hangzhou 310018

[2] School of management, Shanghai University, Shanghai 200444

Abstract: [**Purpose/significance**] Technology opportunity prediction helps national and enterprise managers to identify the future direction of technology development, to adjust the development strategy and occupy a favorable posture for technology competition. [**Method/process**] In this paper, a deep learning-based technology opportunity prediction method was proposed. Firstly, the Affinity Propagation (AP) clustering algorithm was applied to achieve the subject classification of technology fields. Secondly, the Doc2Vec algorithm was used to calculate the similarity of patent texts in each technology area, and then identify the technology areas with development potential. Thirdly, the Generative Topographic Mapping (GTM) algorithm was used to map the patent areas with development potential, and we got the technology opportunities through GTM inverse mapping. Finally, a link prediction model based on deep learning was constructed to predict the links of the identified technology opportunities, to obtain the technology opportunities with high development probability. [**Result/conclusion**] This paper uses new energy vehicle patent data to verify the effectiveness of the method. The results show that the prediction accuracy, recall and F1 value of the deep learning-based link prediction model outperform other prediction models and predict the technology opportunities for new energy vehicles.

Keywords: technology opportunity prediction deep learning link prediction the generative topographic mapping

chinaXiv:202304.00471v1